This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Real-Time Melt Pool Homogenization Through Geometry-Informed Control in Laser Powder Bed Fusion Using Reinforcement Learning

Bumsoo Park, Alvin Chen, and Sandipan Mishra

***Abstract*— This paper presents a real-time geometry-informed control strategy to homogenize melt pool measurements in laser powder bed fusion (L-PBF) using reinforcement learning. The learning control strategy incorporates geometric information of the scan path as well as in-situ melt pool measurements to compute the laser power signal for reducing in-process melt pool inhomogeneities. First, we design and validate a data-driven model to train the reinforcement learning agent in simulation, with the goal of reducing the amount of experimental data needed for training. Using this simulation-based training approach has the added benefit of avoiding unsafe or infeasible experiments, an issue that is often encountered in training the reinforcement learning agent. After training, the learned control strategy attenuates the 1-norm error by 37% and standard deviation by 39% in simulation. We then deploy this learned control strategy in an experimental test bed for a new scan geometry. In this test scenario, the policy achieves a 30% reduction in error, and a 36% reduction in melt pool signal variation, thereby illustrating the potential of reinforcement learning in real-time geometry-agnostic control for L-PBF. Finally, we demonstrate that the reinforcement learning agent delivers the same level of performance as a model-based feedforward controller with PID feedback, with 20× less computational time for a single geometry.**

***Note to Practitioners*—This work was motivated by the need to develop a practical control algorithm for L-PBF systems. Because L-PBF systems manufacture customized on-demand geometries, it is critical that the control strategy is extendable to and easily optimized for each geometry. Specifically, this effort develops an efficient and robust reinforcement learning control algorithm that can be used across novel part geometries, once trained. The control strategy is designed using a simulation-to-real approach, which is key for avoiding extensive training effort and avoids unsafe training experiments.**

***Index Terms*— Laser powder bed fusion (L-PBF), reinforcement learning, sim-to-real learning, data-driven model, metal additive manufacturing.**

I. Introduction

ADDITIVE manufacturing (AM) technologies have developed considerably over the past decade [1]. Particularly, metal AM processes such as laser powder bed fusion (L-PBF) have seen increasingly widespread adoption in numerous industrial applications such as aerospace, automotive, and medical fields [2], [3], [4]. One key challenge that the technology faces, however, is the quality of the produced parts: the process is prone to variability that results in defects, which can adversely affect the mechanical properties and usability of these parts. As a result, much of current research in metal AM focuses on addressing various aspects of this quality control problem [3], [5], [6].

Among these studies, it is well accepted that homogeneous melt pool properties during the process are desirable over fluctuating melt pools [7], [8], [9], [10], [11], for superior part quality. Accordingly, the control problem in L-PBF is commonly formulated as the regulation of melt pool behavior, to reduce undesirable outcomes such as dross [12] or pore [13] formation. The deviations in melt pool behavior stem from two main factors: *process*-related – process noise (e.g. spatter) and local overheating (due to uneven powder), or *process parameter*-related – geometric features of the scan (e.g. sharp corners, overhang) and process parameters (e.g. material type or scan parameters) [14], [15].

The compensation of these effects are either done through *reactive* or *predictive* approaches, depending on the type of deviation. Reactive approaches typically compensate for layer-wise/in-layer deviations of the melt pool indicator measurements using feedback strategies. Layer-wise control has been investigated in the context of layer-wise feedback algorithms [8], [16], iterative learning control (ILC) [17], [18], [19], or predictive models (that correlate the measurements to surface roughness) to correct the process parameters for the subsequent layer [20], [21]. Real-time feedback control has been investigated to a lesser extent due to the high demand on controller response time (typically 2-5kHz), and the few current studies employ simple feedback algorithms such as PID [22], [23], [24]. On the other hand, predictive measures compensate for geometry (and other process-parameter)-related effects in an a-priori manner. The majority of these studies use physics-based or empirical models that can predict geometry(process-parameter)-dependent behavior of the melt-pool [14], [25], [26]; and use these models to determine appropriate laser power profiles and scan paths to accomplish a desired process outcome through feedforward control [27], [28], [29], [30], [31], [32], [33]. These studies altogether demonstrate the efficacy and feasibility of the control strategies

Manuscript received 20 November 2023; revised 6 February 2024; accepted 6 April 2024. This article was recommended for publication by Associate Editor H.-J. Kim and Editor J. Li upon evaluation of the reviewers' comments. This work was supported by NSF Civil, Mechanical and Manufacturing Innovation (CMMI) under Award 2222250. *(Corresponding author: Bumsoo Park.)*

The authors are with the Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: parkb5@rpi.edu; chena17@rpi.edu; mishrs2@rpi.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2024.3386882.

Digital Object Identifier 10.1109/TASE.2024.3386882

1545-5955 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

for each aspect respectively but do not combine reactive and predictive capabilities simultaneously.

Although a combination of the aforementioned feedforward and feedback algorithms may at first glance appear a viable option, solving an optimization problem for *every* geometry (i.e., for each layer or scan path) proves to be computationally laborious and expensive. Further, parameter design/tuning for the feedback algorithms requires substantial engineer time. In practical application, the vast majority of AM-produced parts have diverse cross-sectional geometries across different layers, along with differing rastering strategies. Furthermore, the complex nature of L-PBF [14] requires conventional optimization-based methods to employ substantial model order reductions. Thus, designing a model-based control strategy with both reactive and predictive capabilities, without the need for substantial hand-tuning for different process parameters, is challenging.

Data-driven algorithms such as reinforcement learning (RL) [34] can be used to derive an appropriate control strategy, because of their ability to directly determine a control strategy from the input-output relationship of the system. RL algorithms construct control strategies (so-called policies) through a trial-and-error process, by repeatedly observing the effect of actions on output of the system and updating the control strategy until an optimal control strategy is achieved. In model-free RL, this control strategy is developed without a formal analytical model of the system dynamics or the need for direct model inversion through optimization. Additionally, RL algorithms, once trained, show robust capabilities in dynamically changing conditions by learning the latent properties of the system from data, which can improve the applicability of the trained RL agent for varying scan paths or geometries. These characteristics together make RL a strong candidate in the development of an effective control strategy for L-PBF.

Despite this potential, the development of RL-based control strategies in L-PBF systems has challenges. First, the search for the optimal control strategy during the learning phase requires considerable exploration (trial-and-error) prior to convergence, resulting in extensive training periods and more importantly, posing safety issues (e.g. abnormally high/low laser power, laser power oscillations) during experiments. It is therefore not feasible to directly learn optimal control strategies through experiments on the physical system. Second, the performance of RL-based algorithms heavily relies on appropriate feature selection and reward design [35], [36], [37], [38], requiring meticulous commissioning and design effort.

A. Related Work and State-of-the-Art

The application of RL for L-PBF control is a relatively recent concept and thus has not been investigated widely, compared to other machine learning methods [39]. While there are several studies that use RL in other metal AM processes (e.g. wire arc AM [40] or laser welding [41]), there has been limited application of RL algorithms for L-PBF control/process optimization [21], [42]. In Ogoke and Farimani [42], deep reinforcement learning (DRL) is used to determine appropriate laser power and velocity values to control the melt depth, by using consecutive images of the cross-sectional melt pool heat maps in a simulation environment. The cross-sectional temperature field images are provided to the controller, in which the controller then directly returns a corresponding laser power and speed for the next timestep. Knaak et al. [21], proposed the usage of model-based RL (MBRL) for layer-wise process parameter optimization (laser power and scan velocity) to reduce deviations in the surface roughness and defective regions (characterized by a separate model). High dynamic range (HDR) images are used to train a model to determine the surface roughness and defects, which is used as the state information for the MBRL controller. The authors showed that the algorithm was able to effectively reduce the surface roughness. While both studies demonstrate an effective use of RL for L-PBF process control, the knowledge gap remains regarding: (1) a control algorithm that exhibits predictive and reactive capabilities, without requiring extensive tuning or optimization efforts, (2) the experimental demonstration of a *real-time* learning-control approach, and (3) the demonstration of a safely and timely developed a learning-based approach in L-PBF.

To address this, in [43], we demonstrated a geometry-informed RL-based approach, where the control strategy is derived using RL in a simulation environment. This previous study showed that the RL agent can be trained with minimal effort and was exportable to novel geometries. However, since noise and other unmodeled physics were not captured in the simulation, the agent's predictive–reactive capabilities were not tested to their full extent for experimental demonstration. Building upon [43], in this paper we experimentally demonstrate a strategy inspired by the simulation-to-reality (sim-to-real) approach employed in various control applications [44], by deploying the simulation-trained control strategy from [43] in the physical system to resolve the issue of *safely* training the algorithm. Next, to demonstrate the predictive–reactive capabilities of the control algorithm, we experimentally evaluate the real-time control algorithm. Through the experimental validation, we find that the simulation-trained policy (1) demonstrates both predictive and reactive capabilities when deployed on the physical system, (2) does not require re-optimization upon deployment in the physical system, and (3) can homogenize the melt pool measurements, which ultimately leads to improved build part quality.

The core contributions of this work (in contrast to [43]) can therefore be summarized as:

1) Safe training and implementation of a learning-based control strategy for an experimental L-PBF system using a sim-to-real deployment approach.
2) Development and demonstration of a real-time RL-based control strategy that incorporates geometric information (in a predictive manner) along with feedback control in L-PBF, and that is applicable to novel geometries/scan-paths in the physical system without any further tuning or modification once learned.
3) Experimental validation of the developed *real-time* control strategy in a physical L-PBF system to show

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

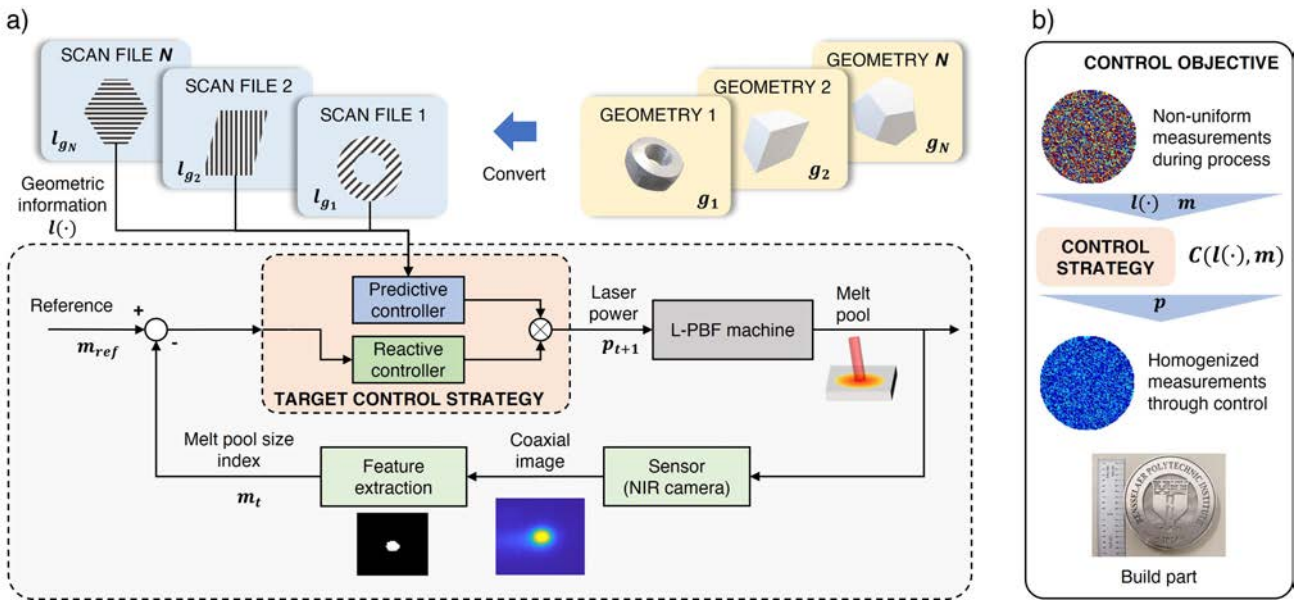

Fig. 1. Geometry-agnostic control melt pool homogenization problem for L-PBF. (a) The goal of this work is to find a suitable control strategy, that can *anticipate* geometric effects and *respond* to in-layer deviations in melt pool measurements, without having to solve an optimization for every available geometry. (b) The objective of the controller is to appropriately harness the given geometry information $l(\cdot)$ and measurement information m, to yield a proper laser power value for the next timestep, ultimately to homogenize the melt pool measurements.

reduction in melt pool deviation and error, demonstrating an improvement in build quality; benchmarking against a conventional feedforward+PID controller.

II. Problem Description

We first provide a description of the experimental system used in this study, followed by a formal statement of the L-PBF control problem.

A. Experimental Setup

The model identification and experimental validation in this work are based on the open architecture L-PBF system presented in [17] (Fig. 2). This system can build parts up to 50mm × 50mm in cross-sectional size, with commercially available metal powders such as stainless steel. The PBF machine is equipped with a $400W$ NdYAG laser, a SCANLAB intelli$SCAN_{de}$20 galvoscanner for the actuation, and a coaxial NIR (Near-IR) camera setup (similar to [8], [22], and [45]) to monitor the melt-pool during the process. A Basler acA2000-165umNIR camera is used to acquire coaxial images of the melt pool in the near-IR band (800-950 nm) at 2kHz. All images are formatted as 8-bit intensity images, with a size of 64 × 64 pixels in size, yielding an instantaneous field of view of 22 μm per pixel.

Low-level control of the machine (e.g. laser firing and positioning) is handled by a SCANLAB RTC5 control board, while high-level control is accomplished using America Makes software [46], supplemented with custom C++ code developed to provide auxiliary functionality. The process is initiated by reading in a scan file, consisting of a list of straight lines, default power value for each line, and scan speed. For an open-loop scenario, each scan line would be executed line-by-line, with the specified speed and default power value. Note that the default power values can be overwritten during the control loop, and thus the control algorithm is able to compute and apply a new power value depending on the observation. The image frame acquisition and power command (computation) are synchronized, i.e., the power values of the laser are updated after every image acquisition ($500\mu s$). Note that, based on our previous timing studies, the time jitter of the image acquisition was found to be 15-30 μs.

B. L-PBF Control Problem Formulation

Fig. 1 illustrates the geometry-agnostic melt pool homogenization problem addressed in this paper. Based on the system described above, the objective of this research is to develop a control strategy with both reactive and predictive capabilities, without having to optimize for every geometry. As discussed in Section I, because geometric effects have to be dealt with in an a-priori manner, a feedforward controller would typically obtain the laser power profile by solving an optimization problem. This can result in extensive development (optimization) times, especially with part geometries with a large number of varying layer-wise scan paths. Reactive controllers (i.e., feedback algorithms such as PID) require empirical tuning of the gains, which can be time consuming and potentially unsafe (e.g., inappropriately designed control gains may result in abnormal power values that can damage the part and the machine).

Here, our goal is to directly derive the predictive and reactive control law $C(l_{g_n}(\cdot), m)$, which takes in *geometric information* $l_{g_n}(\cdot)$ for an arbitrary part geometry g_n along with real-time melt pool measurement information m, to effectively drive subsequent measurements towards a reference value and reliably homogenize future measurements, i.e., find the mapping

$$C(l_{g_n}(\cdot), m_t) = p^*_{t+1}, \tag{1}$$

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

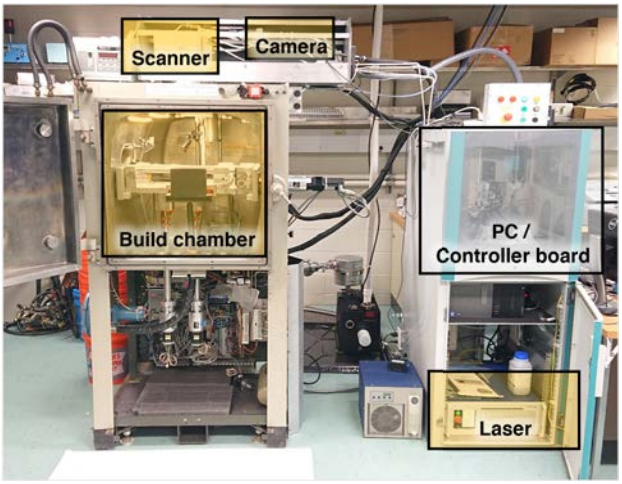

Fig. 2. System description of L-PBF system incorporated in this study. Laser/scanner is used for the actuation, and near-IR (NIR) camera is used to acquire coaxial melt pool images at a rate of 2kHz.

where

$$p^*_{t+1} = \arg\min_{p_{t+1}} \sum_{t=0}^{T-1} ||m_{t+1}(m_t, l_{g_n}, p_{t+1}) - m_{ref}||^2,$$
$$\text{s.t.} \quad p_{t+1} \leq p_{max},$$
$$p_{t+1} \geq p_{min} \ \forall t. \tag{2}$$

Here, $t = 0, 1, \ldots, T-1$ is the timestep of each point along a given scan layer, p_{t+1} is the power value at the next timestep $t+1$, subject to lower p_{min} and upper p_{max} power limits. m_{t+1} is the measurement at time $t+1$ and m_{ref} is the target measurement value. We note here that the signal $l_{g_n}(\cdot)$ must be constructed carefully for a proper representation/interpretation of the geometry, described further in Sections IV and V.

III. Proposed Control Design Methodology

The design methodology consists of three stages, as illustrated in Fig. 3. In Stage 1, a data-driven spatio-temporal model is constructed and identified from experimental data. Next in Stage 2, the RL-algorithm interacts with the model from Stage 1 to learn an optimal control strategy. Finally in Stage 3, the learned control strategy is deployed in the L-PBF machine.

Stage 1: To develop a data-driven model that replicates geometric effects of the process (e.g. overheating during acute turnarounds) while incorporating the sensor dynamics, NIR image data from the process is first collected from the process. Features indicative of the melt pool size are extracted from the images for real-time control, to represent the data as a time series, which are then spatially registered based on the nominal scan path and sampling rate. The spatially mapped data is then used to identify the parameters of a physics-inspired model, providing an environment for the RL training.

Remark: While it is possible to replace Stage 1 (spatio-temporal model development) with a different model (e.g. high-fidelity simulation models), this model would inherently be computationally expensive and time-consuming. More importantly, the discrepancy between the output of a simulation model (typically temperature values; obtained by solving a differential equation) and the actual measurements obtained from the experimental system (melt-pool size index, obtained by extracting features from NIR images) requires incorporation of this sensor behavior.

Stage 2: Next, the model identified in Stage 1 is considered as a black-box environment, with which the RL algorithm interacts to learn the optimal policy. Prior to training, the system inputs and outputs are formulated, along with a suitable reward function to guide the RL policy to achieve a desirable goal (i.e., a prescribed melt pool measurement reference).

Stage 3: In the final stage, the learned policy is transferred and deployed in the physical system for control. To enable real-time control, the NIR images are converted into the information that the RL algorithm can interpret as the images are acquired, while the RL actions are directly converted into laser power values and applied to the machine. In this stage, the trained policy is deployed in an unforeseen test geometry, to demonstrate the geometry-agnostic capabilities of the policy.

IV. Physics-Inspired Model Construction

First, we describe the gray-box model used in the training of the RL algorithm (Stage 1). A low-order model structure was chosen because of the data-hungry nature of RL, i.e., a large number of trials ($10^4 - 10^6$) are often needed to learn an optimal policy. Further, this model was designed to incorporate the sensor behavior in the physical system.

A. Spatio-Temporal Registration

An overview of the spatio-temporal registration process is presented in Figure 4. The first step of the model construction was to extract features from the NIR images captured by the camera. To enable a single-input single-output (SISO) representation of the process for real-time control, we extracted a signature indicative of the melt pool size from the images. We denote this feature (melt pool size index) at time t as m_t.

The temporal signal of m_t was then spatially mapped based on the nominal scan path, assuming ideal trajectory tracking and constant velocity of the scanner. Thus the scan path was interpolated based on the scan speed and sampling time ($\frac{1}{f_s}$, f_s =camera sample rate) to estimate the individual location of m_t. This transformation allowed m_t and $m_i = m(x_t, y_t)$ to be used interchangeably in the spatial map. An example of the resulting transformation is visualized in the lowest block of Fig. 4.

B. Data-Driven Spatio-Temporal Model

The spatio-temporally registered data was then collected to identify the data-driven model. To design a suitable model structure, an autoregressive model inspired by [47] was chosen, which is based on the analytical temperature solution of the heat equation, assuming a Gaussian heat source moving over an infinite plate at constant velocity. The resulting

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

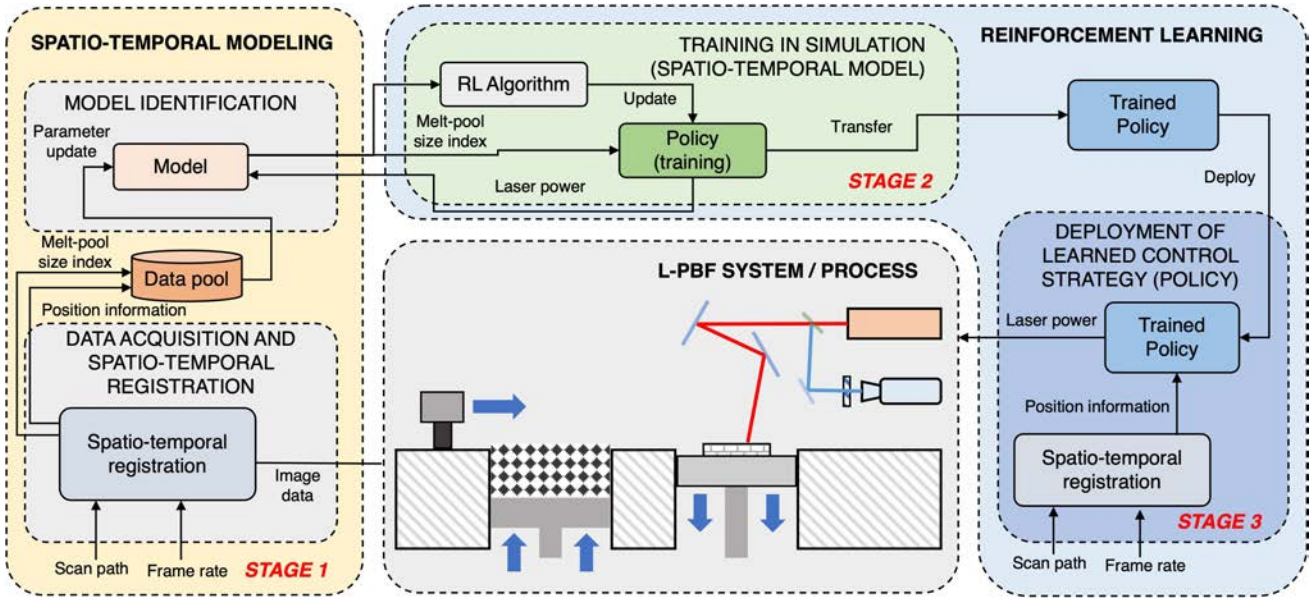

Fig. 3. Overview of the control design strategy. A data-driven spatio-temporal model is first constructed from actual process data (Stage 1). The model is used as a simulator in which the RL controller (policy) is trained (Stage 2)). The trained policy is then deployed to the physical machine, yielding a geometry agnostic feedback controller that can be used for various geometries (Stage 3).

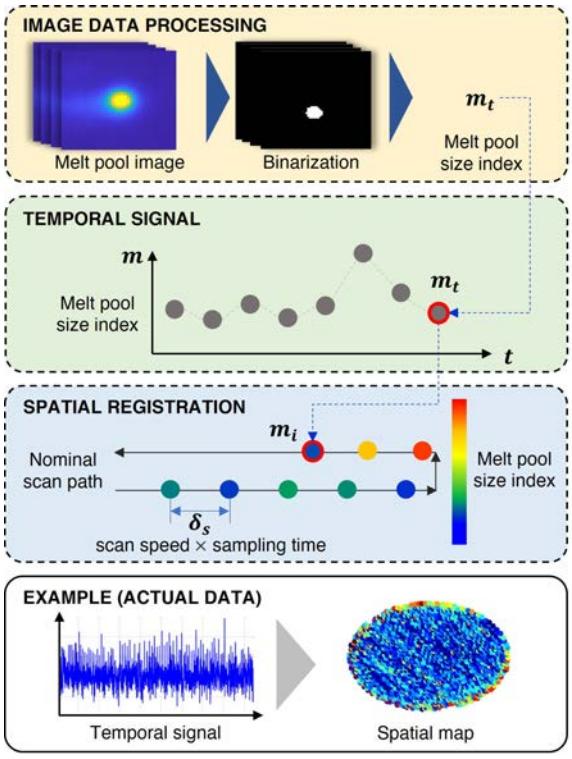

Fig. 4. Spatio-temporal registration process. The stream of images are represented as a time signal through feature extraction (melt pool size index). A spatial map of the melt pool size index values is then obtained by interpolating the nominal scan path, allowing a spatio-temporal transformation of m_t to m_i.

temperature field can be seen as a Gaussian heat source that exponentially decays over time, represented as

$$\hat{m}_{t+1} = \sum_{j=0}^{M} (m_{t-j} \cdot e^{-\lambda_d \Delta d_{tj}{}^2} \cdot e^{-\lambda_t \Delta t_{tj}}) + f(p_t, v), \quad (3)$$

where, $\hat{m}_{t+1}$ is the prediction at time $t+1$. The measurement at time t and $M-1$ previous measurements were considered along the scan path, where each measurement was regarded as a point source with decaying effect over distance and time. Δd_{tj} denotes the distance and Δt_{tj} denotes the difference in time between the current and j^{th} points. The model parameters λ_d and λ_t were identified from experimental data (from Fig. 5) by minimizing mean square prediction error. $f(p_t, v)$ was designed as a function that maps the laser power and speed to a relative effect on the subsequent measurement, which was identified through linear regression of experimental data from a ramp input test.[1]

Once the model structure was defined and the parameters were identified, we evaluated the model by comparing the model predictions with experimental data. Fig. 5 shows a spatial map of the model prediction compared with experimental data. The experimental data was averaged over 10 identical layers to reduce the effects of noise. Here we notice that the model was able to replicate the overheating effects around the turnarounds, with a similar measurement level (mean absolute error of 5.7 and mean percentage error of 1.7%) corresponding to a laser power value.

V. Reinforcement Learning

Next, we recall the preliminary basics for RL design briefly. In RL, the learner is referred to as the *agent*, which learns by interacting with its surroundings, also known as the *environment*, through trial and error (Fig. 6). This process is repeated until a desired control strategy is achieved.

To formulate the RL problem, the system is represented as a Markov Decision Process (MDP), which consists of states

[1]Although both laser power and speed have an effect on subsequent measurements, laser scan speed was kept constant during the model identification due to the inability to modify the scan speed in real-time for the system used in this work.

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

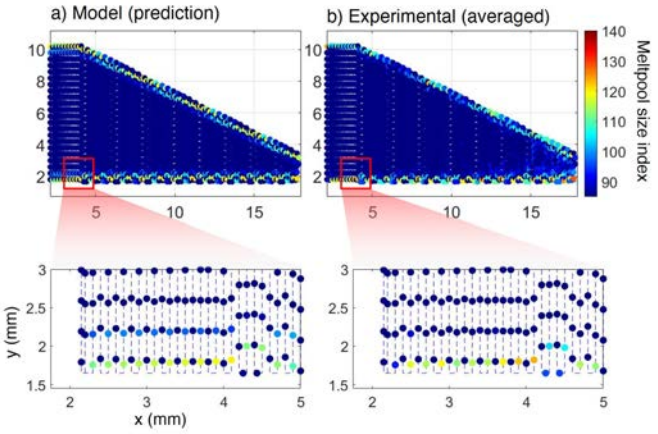

Fig. 5. Model validation. A comparison of the model predictions are shown against the experimental data (averaged over 10 layers, due to high noise levels). Effects such as overheating at turnaround points are represented by the spatio-temporal model in the exact locations.

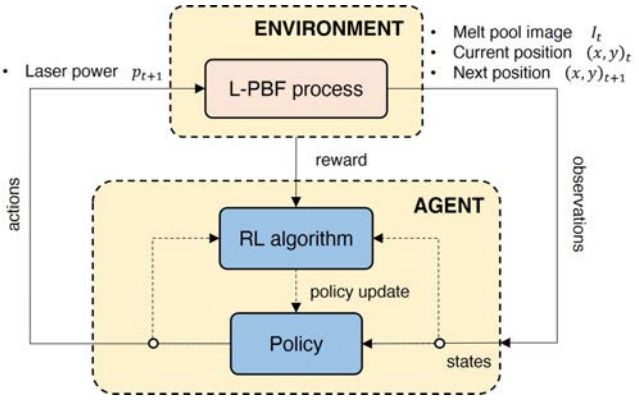

Fig. 6. Reinforcement learning (RL) process the the context of L-PBF control. The goal of RL is to train an agent such that it learns a control strategy (policy) to perform a set of desired actions. The agent learns through repeated interaction with the environment. During each cycle, the agent first takes an action and obtains the subsequent observations and rewards from the environment. The actions, observations, and rewards are then used to update the policy towards a direction that maximizes the expectation of cumulative future rewards.

$S_t \in \mathcal{S}$, actions $A_t \in \mathcal{A}$, rewards $R_t \in \mathcal{R}$, and transition probabilities $Pr(S_{t+1}|S_t, A_t)$. $\mathcal{S}$ is the set of possible observations (measurements), $\mathcal{A}$ is the set of feasible actions, $\mathcal{R}$ is a set of scalar values assessing the current circumstance, and $P(S_{t+1}|S_t, A_t)$ is the probability of transitioning from S_t to S_{t+1} taking action A_t. In a model-free learning scheme, the transition probabilities are unknown. Based on the MDP, the control strategy, known as the *policy* π, is a function that maps states $\mathcal{S}$ to actions $\mathcal{A}$, i.e., $\pi : \mathcal{S} \rightarrow \mathcal{A}$. The objective in RL is to find a policy that maximizes the cumulative future rewards $G_t = \sum_{k=t+1}^{\tau} \gamma^k r_k$, where $\gamma \in [0, 1]$ is the discount factor on future rewards, r is the reward value.

For feasibility of real-time implementation of the algorithm, in this paper we use a value-based algorithm (Q-learning [48]) in its tabular form. Value-based algorithms [34] find the optimal policy by learning a *value function*, defined as the expectation of G_t, i.e., $\mathbb{E}_\pi[G_t]$. The value function with respect to a given state-action pair (s, a), $Q^\pi(s, a)$

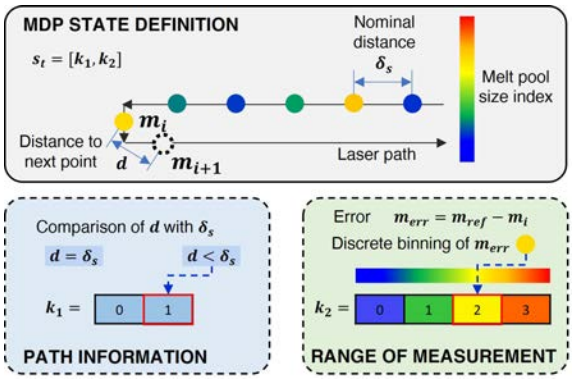

Fig. 7. State definition for MDP. k_1 represents the range of m_{err}, and k_2 represents the proximity of the subsequent point to the current point. The state s_t is represented as a Cartesian product of the two parameters, i.e., $s = [k_1, k_2]$.

(state-action value function), which is written as

$$Q^\pi(s, a) \equiv \mathbb{E}_\pi[G_t \mid S_t = s, A_t = a]. \tag{4}$$

In tabular Q-learning, this value function $Q(s, a)$, is represented in matrix from, where the rows and columns represent the discrete and finite state/action spaces. Therefore both the states and actions of the RL must be formulated accordingly, as discussed below.

A. MDP Formulation

1) State Definition: The discrete states were designed to capture two pieces of information: the value of the current measurement to predict its effect on the subsequent point, and the path information (Fig. 7). We accomplish this by assigning two state elements k_1 and k_2 to represent each piece of information. $k_1 \in \{0, 1\}$ was defined as the path information element, in which determines whether an upcoming point is located at an acute turnaround. This was accomplished by evaluating the Euclidean distance to the subsequent point (denoted by $d = \|m(x_t, y_t) - m(x_{t+1}, y_{t+1})\|$). Because for all points along a straight path d is equal to the sampling distance δ_s (defined as scan speed / sampling rate), any point such that $d < \delta_s$ was considered to have an upcoming turnaround (non-straight path). Thus k_1 was assigned a value of 1 for $d < \delta_s$ and 0 for all other nominal cases.

$$k_1 = \begin{cases} 1 & \text{if } d < \delta_s, \\ 0 & \text{if } d = \delta_s. \end{cases} \tag{5}$$

Next, the measurement-related state element $k_2 \in \{0, 1, 2, 3\}$ was assigned a value based on discrete binning of the error $m_{err} = m_{ref} - m_i$. A total of 4 discrete bins were used, in which the thresholds for each bin were determined heuristically from experimental data. The number of bins were empirically determined to represent the cold, nominal, slightly hot (due to slight heat accumulation), and hot measurements (due to turnarounds and extreme heat accumulation), respectively. The final state s_t was then represented as the Cartesian product of the two parameters, i.e., $s = [k_1, k_2]$, resulting in a total of 8 possible states.

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

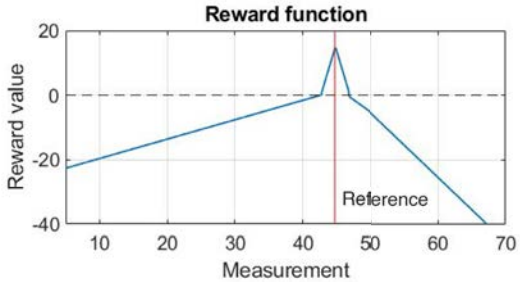

Fig. 8. Visualization of reward function with respect to a given example reference $m_{ref} = 44$. Constructed as a piece-wise linear function with varying slopes, to provide higher rewards for smaller errors. Asymmetric design was chosen to discourage overheating more than undermelting.

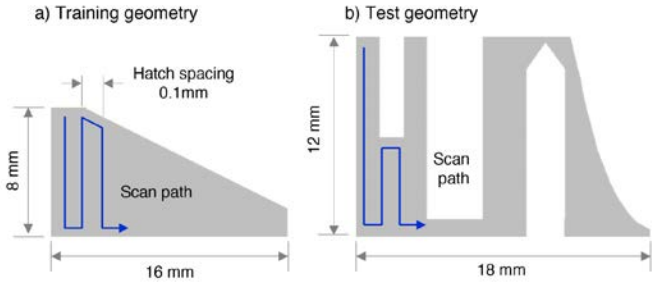

Fig. 9. Part geometry used for RL training and testing. (a) A triangular geometry was used, with hatch spacing set as 0.1 mm. (b) A relatively more complex geometry was used for deployment performance evaluation and experimental validation. Hatch spacing and scan pattern was identical to that of the train geometry.

2) Action Definition: To discretize the action space, we defined a set of discrete power values centered around the constant open loop power value p_{OL}. A total of 16 values ranging from $p_{OL} - 50$ to $p_{OL} + 30$ with increments of $5W$ were chosen. The increment value was heuristically determined as the minimum value to produce a perceptible change through melt pool measurement variance, yet without excessively expanding the action space. Unlike function approximation methods (e.g. DQN, Policy gradients), tabular Q-learning does not explicitly define the similarity between states. Hence each state-action pair is treated independently in the Q-function, and the algorithm relies on exploration to encounter and update the Q-values for different state-action pairs.

3) Reward Construction: Finally, we constructed the reward function to guide the policy towards a strategy to minimize error, in which an analogous structure was chosen (Fig. 8). The reward function was designed as piece-wise linear function, with varying slopes to incentivize minimizing the error. These slopes were computed from heuristically assigned reward values at different points, e.g., $+15$ at m_{ref}. Measurements within a certain vicinity of the reference were all rewarded with positive reward values, whereas measurements outside the vicinity were penalized with negative reward values. Note that an asymmetric reward design was chosen to discourage overheating more than undermelting.

Based on the formulation of the MDP, the measurements and actions from the model were transformed into MDP states and actions, respectively. The states s_t, actions a_t, and rewards r_t were correspondingly used to update the value function $Q(s, a)$ through the Q-learning algorithm (6) [48].

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)), \quad (6)$$

where α is the learning rate for the Q function update.

VI. Sim-to-Real Experimental Deployment of Learned Control Strategy

After $Q^*(s, a)$ was found, we transferred $Q^*(s, a)$ to the physical machine (Stage 3). $Q^*(s, a)$ was implemented as a look-up table such that a desirable action is returned for a given state under a greedy policy (7), i.e., the action with the highest Q value was chosen for a given state S_t, i.e., the laser power update is determined by

$$\pi(S_t) = a \equiv \arg\max_{a \in A} Q^*(S_t, a). \quad (7)$$

For real-time implementation, the feature extraction from the NIR images and corresponding state conversion were integrated into the supervisory machine control codes. Note that the spatio-temporal registration only needed to be executed once at the beginning of the layer, hence the spatial map was generated in a layer-wise manner. Similar to the states, the actions a were converted back into corresponding power values, based on the constant open loop power P_{OL}.

VII. Simulation and Experimental Results

The training of the RL algorithm was done with the geometry shown in Fig. 9 (a). Here we first analyzed the training results and the performance of the trained policy within the simulation. We next analyzed the policy and action decisions yielded by the controller and compared with a relatively more complicated RL algorithm to show that the tabular Q-learning can produce comparable results. Finally, we evaluated the performance of the simulation-trained controller when deployed in a physical system, applied to a novel scan geometry to show geometry-agnostic capabilities.

A. Training Results

The RL algorithm was trained for 200 iterations, converging at approximately 100 iterations (Fig. 10). The geometry shown in Fig. 9 (a) was used for the training, and an entire layer was considered a single iteration (episode). We designed the scan path with $0.1mm$ hatch spacing and $800mm/s$ scan speed, and used an open loop power value of $250W$. $P_{OL} = 250W$ was heuristically determined from prior tests to ensure adequate performance. The hyperparameters for the RL algorithm were empirically tuned as the following: the discount factor $\gamma = 0.7$ and the learning rate $\alpha = 0.2$.

1) Note on Computational Time: The training time for the geometry-agnostic RL algorithm was 17 seconds. The training was executed on an Intel i7-9700K CPU with 16 GB of memory. Note that the time required to derive a model-based

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

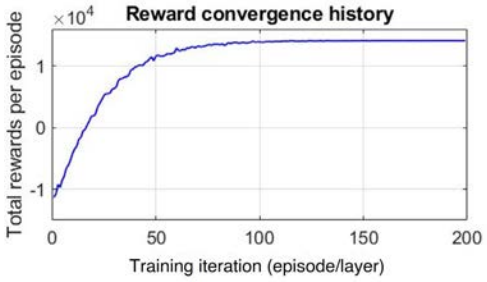

Fig. 10. Reward history during training. RL algorithm was trained for 200 iterations (episode) on the training geometry shown in 9 (a). The cumulative rewards per episode converged at approximately 100 iterations.

feedforward controller for the training geometry (single geometry) was 340 seconds. The optimization of the model-based feedforward was demonstrated to emphasize the need for a geometry-agnostic controller, as the development time would linearly grow with layer number/part-geometry. Experimental results for the model-based feedforward combined with feedback will be discussed in Section VII-D.

We next evaluated the performance of the trained policy within the simulation and analyzed the policy. Here, we visualized the inputs (power profile), resulting melt pool indicator measurements, and 1-norm error with respect to m_{ref} of the open-loop and RL-controlled cases (Fig. 11). The visualization is based on the spatial registration discussed in Section IV-A, and the color of each point represents a proportional value in each category. We notice that the RL-controlled case learned to effectively reduce the power along the edges where the turnarounds occur, and additionally adjusting the power value in adjacent points. Such a strategy resulted in a more uniform measurement map. The overall error and melt pool signal variation[2] were reduced by 37% and 39%, respectively, showing that the RL was able to learn an effective strategy within the simulator.

B. Comparison With Other RL-Algorithms

To ensure that the tabular setup can achieve an optimal policy similar to that of a more complex RL algorithm, we compared the trained policy against a function approximation-based RL algorithm. The algorithm used for comparison was REINFORCE [49], a policy-based algorithm that uses a function approximator to directly learn the policy. We used a neural network with 2 hidden layers, each with 100 nodes for the function approximator. As function approximator methods are capable of handling high-dimensional continuous state-action spaces, instead of discretizing the states, we directly used the normalized values of the measurement m_t and distance to the subsequent point d in the states, i.e., $s_{PG_t} = [m_t, d]$. On the other hand, the same set of actions were used, for a fair comparison.

Spatial maps of the power profiles from each algorithm are compared in Fig. 12. We noticed that both algorithms learned similar strategies, by lowering the laser power values along the edges, albeit the Q-learning being relatively more aggressive

[2]Variation is defined as standard deviation; $\sqrt{\frac{1}{N}\sum |m_i - \mu_m|^2}$.

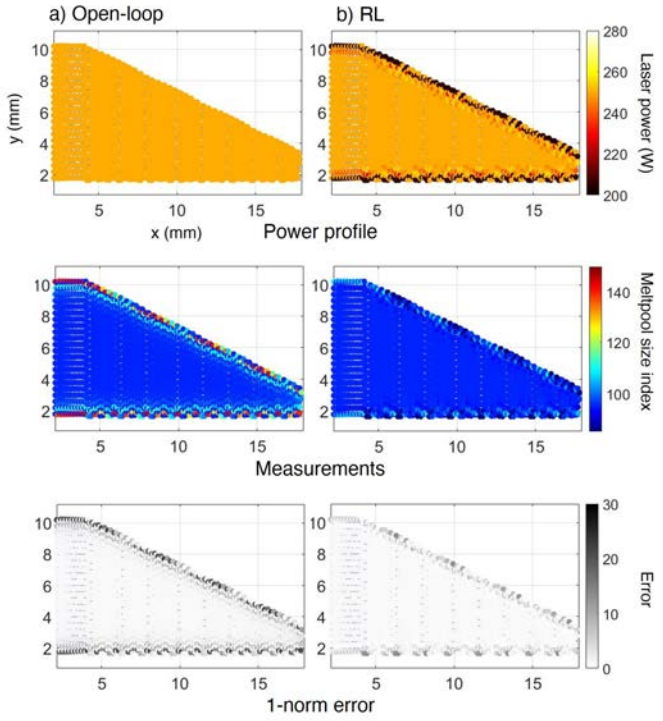

Fig. 11. Performance evaluation in training geometry. Performance of the proposed RL algorithm is compared against open-loop. Spatial maps of the power profiles, measurements, and absolute errors are compared. The reference melt pool signal was set to $m_{ref} = 82$, derived as the mean nominal (in-line value without overheating) value of the open-loop test. The RL reduced melt pool signal variation by 37% and 1-norm error by 39%, compared to the open-loop case.

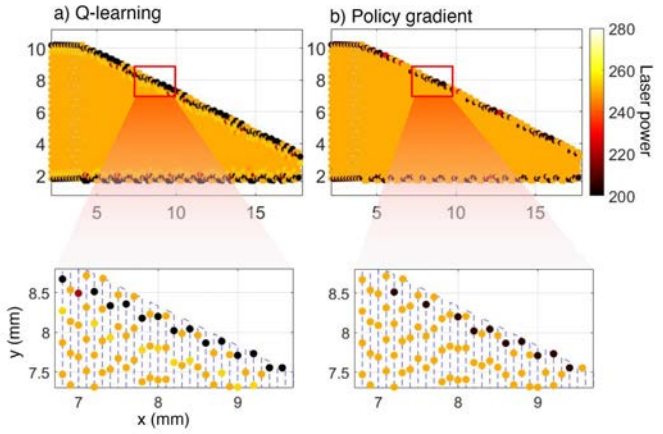

Fig. 12. Policy comparison with another RL algorithm. A function approximation policy-based algorithm (REINFORCE) was used as comparison. Similar policies are derived, both of which lower the power along the edges where the turnarounds occur.

in lowering the power in the edges. For instance, the policy trained with Q-learning lowered power values for all points with $d < \delta_s$, whereas the policy trained with REINFORCE lowered the power only at points where $d \ll \delta_s$. This can be attributed to the fact that the distance information was directly provided to the states in REINFORCE, such that an appropriate distance threshold for lowering the power was heuristically determined. Nonetheless the overall policies were similar in terms of performance, suggesting that the Q-learning

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

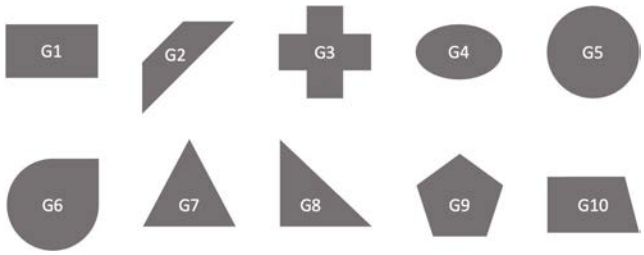

Fig. 13. Part geometries used in geometry-informed capability test. A total of 10 geometries were tested with the same policy, learned in Section VII-A.

TABLE I
SUMMARY OF GEOMETRY-INFORMED CAPABILITY TEST

Part number	Error reduction (%)	Variation reduction (%)
G1	49	61
G2	30	44
G3	50	62
G4	42	57
G5	45	57
G6	49	61
G7	45	56
G8	26	42
G9	36	47
G10	37	50

algorithm can learn an equivalently effective control strategy despite its lower computationally complexity. It is worth noting that the target L-PBF machine currently is not capable of parallel computing, and thus the REINFORCE algorithm cannot be executed in real-time in the target machine. For such reasons, experimental validation was demonstrated for Q-learning only.

C. Validation of Geometry-Informed Capabilities

Prior to the experimental validation, we validated the learned policy with respect to 10 novel geometries (Fig. 13) within the simulator. All scan paths were serpentine paths starting from the left of the geometry and ending at the right. Note that the *same* policy was used for *all* geometries, to show that a single policy is applicable to unknown geometries, as long as the scan path is provided.

The assessment of the performance in each geometry was done with respect to the same performance metrics from Section VII-A, i.e., error and melt pool signal variation reduction values (Table I). For all geometries, the trained policy exhibited noticeable error and signal variation reduction capabilities, demonstrating the geometry-informed capabilities of the RL policy.

D. Experimental Validation of Learned RL Policy

The trained policy was finally deployed in the physical system (Stage 3 of Fig. 3) for the experimental validation. The test geometry shown in Fig. 9 (b) was used. Multiple features from the geometry-informed test set were compiled to design a challenging test geometry, with narrow channels and varying scan-lengths, as opposed to the relatively simple training geometry (Fig. 9 (a)). Scan parameters such as hatch spacing and open-loop laser power were identical to that of the training case. The deployment of the simulation-trained policy in the physical system with respect to the test geometry confirmed the following: (1) the algorithm is applicable to novel scan-paths/geometries without further modification once the policy is learned, and (2) the demonstration shows the feasibility of a simulation-learned control strategy executable in real-time, in a physical system.

As shown for the training case, we visualized the spatial maps of the power profiles, measurements, and 1-norm errors (Fig. 14). For the implementation, due to the higher levels of noise in the actual system, we incorporated an exponential filter to smooth the measurements acquired in real-time. For similar reasons, the visualization is based on averaged values across 10 layers (as demonstrated in Fig. 5). Notably, the simulation-learned policy appropriately adjusted the power values along the edges and narrow regions of the scan part, resulting in a smoother measurement map even in the physical system. More importantly, the controller mitigated local overheating throughout the narrow channels through feedback, as this effect was not adequately captured in the data-driven model. This effect can also be found in the regions with shrinking scan length, i.e., the power is gradually lowered in the -5 to -1mm and 2 to 6mm region in the x-axis (Fig. 14 (c), power profile). This strategy effectively eliminated most of the overheating measurements throughout the narrow channels and edges (where the turnarounds occur).

The same reference value $m_{ref} = 82$ from the simulator was chosen. No further efforts were made to resolve the discrepancy between the simulation and physical-system, as apposed to the majority of literature that investigate sim-to-real training approaches. This is mainly due to the fact that the data-driven model was already identified to replicate the dynamics of the process in terms of setpoints, i.e., the model parameters were chosen such that $\hat{m}_t \approx m_t$ for a given laser power value and scan path, avoiding additional tuning efforts. The experimental validation of the controller showed a 30% reduction in error and 36% reduction in melt pool signal variation in an unforeseen geometry, well demonstrating the geometry-agnostic capabilities of the proposed algorithm.

Benchmarking Against Feedforward-Feedback Control

In addition to the comparison with open loop results, we compared the RL to a conventional feedforward(FF)-feedback(FB) controller (Fig. 14 (b)). The FF controller was derived by framing an optimization problem based on the model developed in Section IV-B, with respect to a given scan geometry, i.e., a layer-wise power profile P^* was derived. Denoting the model as $G_{\lambda_d,\lambda_t}(p_t)$, the optimal power profile was found by solving

$$\begin{aligned} P^* = \arg\min_{\tilde{P}} \quad & \sum_{t=0}^{T-1} ||G_{\lambda_d,\lambda_t}(p_{t+1}) - m_{ref}||^2, \\ \text{s.t.} \quad & p_{t+1} \leq P_{max}, \\ & p_{t+1} \geq P_{min} \quad \forall t. \end{aligned} \tag{8}$$

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

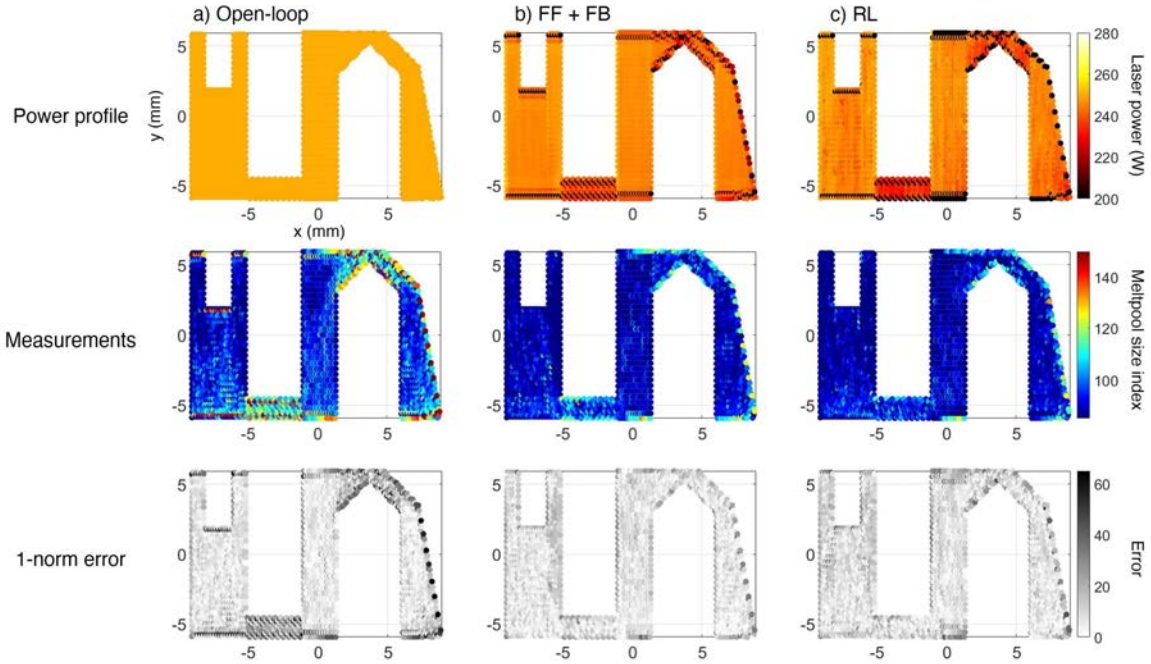

Fig. 14. Experimental validation of performance in novel test geometry. Proposed RL algorithm is compared against open loop and feedforward+feedback control. The feedforward controller was derived through model-based optimization, and an empirically tuned PID was used for the feedback control. Reference melt pool size index value was set as $m_{ref} = 82$, identical to that of the simulation. The same spatial maps (as Fig. 11) are shown for the experimental validation, averaged across 10 layers due to high levels of noise. The RL reduced the absolute error by 30% and the variation by 36%, whereas the feedforward+feedback reduced the error by 32% and the variation by 35%. The RL was able to achieve comparable results to the feedforward+feedback despite having a substantially shorter optimization time.

where T is the total number of timesteps in a scan layer, and $\tilde{P} \triangleq [p_1, p_1, \ldots, p_T]$ is the vector of power values for a scan layer. Since there are no suitable ODE models available for feedback design, the FF controller was combined with an empirically-tuned PID controller. We noticed that the FF+FB yielded similar power profiles to the RL, by lowering power values along the edges and narrow channels. The reduction in error and variation was 32% and 35%, respectively. Although there was no notable difference in performance between the FF+FB and RL, the slightly higher error reduction can be attributed to the fact that continuous power values were used for the FF+FB, allowing a slightly finer adjustment. In contrast, the variation, was further reduced in the RL case. This is mainly due to the slight difference in power reduction strategies along the edges — because the RL lowers the power for *all* points categorized as a turnaround (according to the state definition), the edges tend to overheat less, resulting in a slightly lower variation.

Although both the RL and FF+FB exhibited similar performance, the key advantage of RL lies in the development time (Table II). The model-based feedforward required approximately 340 seconds to optimize for a given geometry in the same hardware used for RL (Section VII-A), implying that build parts with larger layer numbers can require an extensive development time. On the other hand, the RL required only 17 seconds to learn a geometry-agnostic control strategy that is applicable to any geometry. While it would be possible to reduce the optimization time for the FF, the fact that the optimization time for each scan pattern is non-zero, implies that the total time would grow with varying scan paths and geometries. These results altogether, further support the efficacy and feasibility of the proposed method in L-PBF control.

TABLE II
SUMMARY OF EXPERIMENTAL RESULTS

Controller type	Error reduction (%)	Variation reduction (%)	Optimization/Development time (sec)
FF + FB	32	35	340 (per geometry)
RL	30	36	17 (once)

E. Note on Build Quality

Although no direct quantification of the material properties was done in this study, we observed that the homogenization of the measurements resulted in a relatively more uniform surface finish. Fig. 15 shows two example regions of the actual build part, for the open-loop case (Fig. 15(a)) and the RL case (Fig. 15(b)). We can observe a bead-like structure in the open-loop case, which is less noticeable in the RL-controlled case. The higher measurement maps also seemed to have induced higher walls in the narrow channel, in which we find that the RL was able to attenuate the inhomogeneity in surface height as well.

Remark: Note that because the inputs (actions) are bounded (Sec.V), the output (measurements) of the system is expected to be bounded as the system is open-loop stable. Hence the controller is regarded as bounded-input bounded-output (BIBO) stable, and for such reasons, additional stability analysis was omitted from the scope of this study.

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

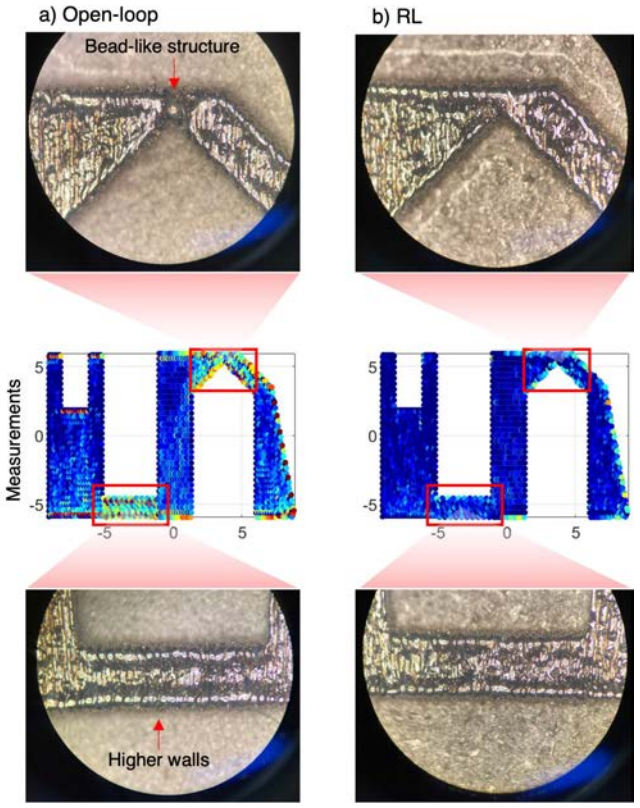

Fig. 15. Qualitative analysis of build quality. (a) Part produced without control. A bead-like structure is observable in the tapered region of the validation geometry. Similar effects are found in the narrow channel. Note that both regions exhibit uneven melt pool measurements. (b) Part produced with RL control. Homogenization of the measurements resulted in a relatively more uniform surface profile. Bead-like structure and higher walls are less noticeable.

VIII. Conclusion

In this study, we developed and tested an RL-control strategy for an L-PBF system that can anticipate geometric effects while responding to in-situ measurements, for the homogenization of melt pool measurements during the process. This is, to our knowledge, the first demonstration of a geometry-agnostic RL-trained control strategy deployed to a physical L-PBF system for real-time control, that is applicable to novel scan-paths and geometries without further tuning or modification. Although being a preliminary study, we have confirmed that the simulation-trained strategy demonstrated a mitigation of training efforts and elimination of safety issues during the development. Moreover, the intermediate results show that from the experimental validation, the simulation-trained algorithm was able to reduce up to 30% in error and 36% in signal variation, well supporting the feasibility of a new approach for geometry-agnostic L-PBF control.

Future work will address the following issues: first, no direct evaluation on the mechanical properties was conducted, as the main scope of this study was focused on a methodology to homogenize the in-situ melt pool measurements. Although there are supporting studies that suggest the correlation between melt pool behavior and mechanical property of the built part, future work will address this issue by directly assessing the improvement in terms of mechanical properties. Second, this work does not address the potential usage of improved hardware in the target (L-PBF) machine, such as the usage of field-programmable gate arrays (FPGA). Usage of advanced hardware can increase the amount of available computational resources, and thus a more complex algorithm with enhanced capabilities can be implemented for the real-time execution. Finally, the SISO representation of the melt pool image through feature extraction exhibited relatively high levels of noise, requiring additional filters to be implemented for the algorithm to run in real-time. Features that are more robust to such fluctuations are of interest and thus will be investigated as a part of future work.

References

[1] O. Abdulhameed, A. Al-Ahmari, W. Ameen, and S. H. Mian, "Additive manufacturing: Challenges, trends, and applications," *Adv. Mech. Eng.*, vol. 11, no. 2, Feb. 2019, Art. no. 168781401882288.

[2] D. Türk et al., "Additive manufacturing with composites for integrated aircraft structures," in *Proc. Int. SAMPE Tech. Conf.*, 2016, pp. 1404–1418.

[3] Y. Zhai, D. A. Lados, and J. L. LaGoy, "Additive manufacturing: Making imagination the major limitation," *JOM*, vol. 66, no. 5, pp. 808–816, May 2014.

[4] C. Y. Yap et al., "Review of selective laser melting: Materials and applications," *Appl. Phys. Rev.*, vol. 2, no. 4, 2015, Art. no. 041101.

[5] J. C. Fox, S. P. Moylan, and B. M. Lane, "Effect of process parameters on the surface roughness of overhanging structures in laser powder bed fusion additive manufacturing," *Proc. CIRP*, vol. 45, pp. 131–134, Jan. 2016.

[6] D. Wang, Y. Yang, Z. Yi, and X. Su, "Research on the fabricating quality optimization of the overhanging surface in SLM process," *Int. J. Adv. Manuf. Technol.*, vol. 65, nos. 9–12, pp. 1471–1484, Apr. 2013.

[7] W. E. King et al., "Laser powder bed fusion additive manufacturing of metals: Physics, computational, and materials challenges," *Appl. Phys. Rev.*, vol. 2, no. 4, 2015, Art. no. 041304. [Online]. Available: https://e-reports-ext.llnl.gov/pdf/800252.pdf http://aip.scitation.org/doi/10.1063/1.4937809

[8] E. Vasileska, A. G. Demir, B. M. Colosimo, and B. Previtali, "A novel paradigm for feedback control in LPBF: layer-wise correction for overhang structures," *Adv. Manuf.*, vol. 10, no. 2, pp. 326–344, Jun. 2022.

[9] J. Wang, R. Zhu, Y. Liu, and L. Zhang, "Understanding melt pool characteristics in laser powder bed fusion: An overview of single- and multi-track melt pools for process optimization," *Adv. Powder Mater.*, vol. 2, no. 4, Oct. 2023, Art. no. 100137.

[10] M. Adnan, H.-C. Yang, T.-H. Kuo, F.-T. Cheng, and H.-C. Tran, "MPI-based system 2 for determining LPBF process control thresholds and parameters," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 6553–6560, Oct. 2021.

[11] A. G. Demir, L. Mazzoleni, L. Caprio, M. Pacher, and B. Previtali, "Complementary use of pulsed and continuous wave emission modes to stabilize melt pool geometry in laser powder bed fusion," *Opt. Laser Technol.*, vol. 113, pp. 15–26, May 2019.

[12] T. Craeghs, S. Clijsters, E. Yasa, F. Bechmann, S. Berumen, and J.-P. Kruth, "Determination of geometrical factors in layerwise laser melting using optical process monitoring," *Opt. Lasers Eng.*, vol. 49, no. 12, pp. 1440–1446, Dec. 2011.

[13] C. Du et al., "Pore defects in laser powder bed fusion: Formation mechanism, control method, and perspectives," *J. Alloys Compounds*, vol. 944, May 2023, Art. no. 169215.

[14] S. R. Narasimharaju et al., "A comprehensive review on laser powder bed fusion of steels: Processing, microstructure, defects and control methods, mechanical properties, current challenges and future trends," *J. Manuf. Processes*, vol. 75, pp. 375–414, Mar. 2022.

[15] B. Lane, S. Moylan, E. P. Whitenton, and L. Ma, "Thermographic measurements of the commercial laser powder bed fusion process at NIST," *Rapid Prototyping J.*, vol. 22, no. 5, pp. 778–787, Aug. 2016.

[16] E. Vasileska, A. G. Demir, B. M. Colosimo, and B. Previtali, "Layer-wise control of selective laser melting by means of inline melt pool area measurements," *J. Laser Appl.*, vol. 32, no. 2, May 2020.

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[17] A. Shkoruta, W. Caynoski, S. Mishra, and S. Rock, "Iterative learning control for power profile shaping in selective laser melting," in *Proc. IEEE 15th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2019, pp. 655–660.

[18] X. Wang, C. S. Lough, D. A. Bristow, R. G. Landers, and E. C. Kinzel, "A Layer-to-layer control-oriented model for selective laser melting," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2020, pp. 481–486.

[19] X. Wang, R. G. Landers, and D. A. Bristow, "Spatial transformation of a Layer-to-Layer control model for selective laser melting," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2022, pp. 2886–2891.

[20] H.-C. Tran, Y.-L. Lo, H.-C. Yang, H.-C. Hsiao, F.-T. Cheng, and T.-H. Kuo, "Intelligent additive manufacturing architecture for enhancing uniformity of surface roughness and mechanical properties of laser powder bed fusion components," *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 4, pp. 1–12, 2004.

[21] C. Knaak, L. Masseling, E. Duong, P. Abels, and A. Gillner, "Improving build quality in laser powder bed fusion using high dynamic range imaging and model-based reinforcement learning," *IEEE Access*, vol. 9, pp. 55214–55231, 2021.

[22] J.-P. Kruth, P. Mercelis, J. Van Vaerenbergh, and T. Craeghs, "Feedback control of selective laser melting," in *Proc. 3rd Int. Conf. Adv. Res. Virtual Rapid Prototyping*. New York, NY, USA: Taylor & Francis, 2007, pp. 521–527.

[23] V. Renken, A. von Freyberg, K. Schünemann, F. Pastors, and A. Fischer, "In-process closed-loop control for stabilising the melt pool temperature in selective laser melting," *Prog. Additive Manuf.*, vol. 4, no. 4, pp. 411–421, 2019, doi: 10.1007/S40964-019-00083-9.

[24] A. Shkoruta, S. Mishra, and S. J. Rock, "Real-time image-based feedback control of laser powder bed fusion," *ASME Lett. Dyn. Syst. Control*, vol. 2, no. 2, Apr. 2022, Art. no. 021001, doi: 10.1115/1.4051588.

[25] M. D. Xames, F. K. Torsha, and F. Sarwar, "A systematic literature review on recent trends of machine learning applications in additive manufacturing," *J. Intell. Manuf.*, vol. 34, no. 6, pp. 2529–2555, Aug. 2023.

[26] D. Mahmoud, M. Magolon, J. Boer, M. A. Elbestawi, and M. G. Mohammadi, "Applications of machine learning in process monitoring and controls of L-PBF additive manufacturing: A review," *Appl. Sci.*, vol. 11, no. 24, p. 11910, Dec. 2021.

[27] H. Yeung, B. M. Lane, M. A. Donmez, J. C. Fox, and J. Neira, "Implementation of advanced laser control strategies for powder bed fusion systems," *Proc. Manuf.*, vol. 26, pp. 871–879, Jan. 2018.

[28] H. Yeung, B. Lane, and J. Fox, "Part geometry and conduction-based laser power control for powder bed fusion additive manufacturing," *Additive Manuf.*, vol. 30, Dec. 2019, Art. no. 100844.

[29] H. Yeung and B. Lane, "A residual heat compensation based scan strategy for powder bed fusion additive manufacturing," *Manuf. Lett.*, vol. 25, pp. 56–59, Aug. 2020, doi: 10.1016/J.MFGLET.2020.07.005.

[30] Q. Wang, P. Michaleris, A. R. Nassar, J. E. Irwin, Y. Ren, and C. B. Stutzman, "Model-based feedforward control of laser powder bed fusion additive manufacturing," *Additive Manuf.*, vol. 31, Jan. 2020, Art. no. 100985.

[31] H. Yeung, Z. Yang, and L. Yan, "A meltpool prediction based scan strategy for powder bed fusion additive manufacturing," *Additive Manuf.*, vol. 35, Oct. 2020, Art. no. 101383.

[32] Y. Ren and Q. Wang, "Gaussian-process based modeling and optimal control of melt-pool geometry in laser powder bed fusion," *J. Intell. Manuf.*, vol. 33, no. 8, pp. 2239–2256, Dec. 2022.

[33] A. Shkoruta, B. Park, and S. Mishra, "An empirical model for feedforward control of laser powder bed fusion," 2022, *arXiv:2201.09978*.

[34] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[35] R. Nian, J. Liu, and B. Huang, "A review on reinforcement learning: Introduction and applications in industrial process control," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106886.

[36] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annu. Rev. Control, Robot., Auto. Syst.*, vol. 2, pp. 253–279, May 2019.

[37] C. Arzate Cruz and T. Igarashi, "A survey on interactive reinforcement learning: Design principles and open challenges," in *Proc. ACM Designing Interact. Syst. Conf.*, Jul. 2020, pp. 1195–1209.

[38] R. L. de Freitas Cunha and L. Chaimowicz, "On the impact of MDP design for reinforcement learning agents in resource management," in *Proc. Intell. Syst., 10th Brazilian Conf. (BRACIS)*. Cham, Switzerland: Springer, Dec. 2021, pp. 79–93.

[39] P. Wang, Y. Yang, and N. S. Moghaddam, "Process modeling in laser powder bed fusion towards defect detection and quality control via machine learning: The state-of-the-art and research challenges," *J. Manuf. Processes*, vol. 73, pp. 961–984, Jan. 2022.

[40] A. G. Dharmawan, Y. Xiong, S. Foong, and G. Song Soh, "A model-based reinforcement learning and correction framework for process control of robotic wire arc additive manufacturing," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 4030–4036.

[41] G. Masinelli, T. Le-Quang, S. Zanoli, K. Wasmer, and S. A. Shevchik, "Adaptive laser welding control: A reinforcement learning approach," *IEEE Access*, vol. 8, pp. 103803–103814, 2020.

[42] F. Ogoke and A. B. Farimani, "Thermal control of laser powder bed fusion using deep reinforcement learning," *Additive Manuf.*, vol. 46, Oct. 2021, Art. no. 102033.

[43] B. Park and S. Mishra, "Geometry-agnostic melt-pool homogenization of laser powder bed fusion through reinforcement learning," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics (AIM)*, Jun. 2023, pp. 1014–1019, doi: 10.1109/AIM46323.2023.10196239.

[44] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: A survey," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, 2020, pp. 737–744.

[45] A. Gökhan Demir, C. De Giorgi, and B. Previtali, "Design and implementation of a multisensor coaxial monitoring system with correction strategies for selective laser melting of a maraging steel," *J. Manuf. Sci. Eng.*, vol. 140, no. 4, Apr. 2018.

[46] America Makes. *4039 Development & Demonstration of Open-Source Protocols for Powder Bed Fusion AM*. Accessed: Feb. 2, 2023. [Online]. Available: https://www.americamakes.us/projects/4039-development-demonstration-opensource-protocols-powder-bed-fusion-additive-manufacturing-pbfam

[47] T. Eagar et al., "Temperature fields produced by traveling distributed heat sources," *Weld. J.*, vol. 62, no. 12, pp. 346–355, 1983.

[48] C. J. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 92–279, Jun. 1992.

[49] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 229–256, May 1992.

Bumsoo Park received the B.S. and M.S. degrees in mechanical engineering from the Ulsan National Institute of Science and Technology (UNIST). He is currently pursuing the Ph.D. degree in mechanical engineering with the Rensselaer Polytechnic Institute (RPI), Troy, NY, USA. His research interests include the development of learning control algorithms for mechanical systems with complex dynamics.

Alvin Chen received the B.S. degree in mechanical engineering from Rutgers University. He is currently pursuing the Ph.D. degree with the Rensselaer Polytechnic Institute. He is with the Intelligent Structural Systems Laboratory (ISSL) and the Intelligent Systems, Automation and Control Laboratory (ISAaC), where he combines elements of non-destructive evaluation with advanced additive manufacturing processes to detect faults in situ for metal AM. His current areas of interests include additive manufacturing, process control, and control systems.

Sandipan Mishra received the B.Tech. degree in mechanical engineering from Indian Institute of Technology Madras in 2002 and the Ph.D. degree in mechanical engineering from the University of California at Berkeley in 2008. He is currently a Professor with the Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. His research expertise spans the general area of dynamical systems, autonomy, and control with a particular interest in learning control, optimal control, and precision mechatronics, with applications to autonomous vehicles, advanced manufacturing, and smart building systems.

Authorized licensed use limited to: The Libraries at Rensselaer Polytechnic Institute. Downloaded on January 17,2025 at 18:13:18 UTC from IEEE Xplore. Restrictions apply.